The *y1* Gene of Maize Codes for Phytoene Synthase

Brent Buckner,* Phillip San Miguel,† Diane Janick-Buckner* and Jeffrey L. Bennetzen†

**Division of Science, Northeast Missouri State University, Kirksville, Missouri 63501 and †Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907*

Manuscript received August 30, 1995
Accepted for publication January 22, 1996

ABSTRACT

The cloned *y1* locus of maize was sequenced and found to encode phytoene synthase. Different "wild-type" alleles of the locus were found to differ by the insertion of transposable elements in their promoter and polyA addition regions, and by the length of a CCA tandem repeat series, without any obvious effect on function of the gene. A dominant *Y1* ("wild-type") allele was observed to be expressed at highest levels in the seedling but also in the embryo and endosperm. The *Mu3* transposable element insertion responsible for a pastel allele of *y1*, which gives lowered levels of carotenoids in the endosperm of kernels and seedlings grown at high temperatures, was located in the 5′ end of the gene. Although the size of the transcript from this *y1* mutation suggests that the *Mu3* element provides the promoter for this allele, leaf tissue in this mutant line contained approximately normal amounts of *y1* mRNA. A recessive allele of *y1*, which conditions normal levels of carotenoids in the embryo and seedling, but almost no carotenoids in the endosperm, was found to accumulate normal amounts of *y1* mRNA in the seedling and embryo, while *y1* transcripts were not detected in the endosperm.

CAROTENOIDS constitute one of the most widespread groups of pigments found in nature (reviewed by GOODWIN 1971). In photosynthetic organisms, they function to prevent photooxidative damage of chlorophyll, and presumably other molecules, and act as accessory pigments for absorbing visible light. Plants can also contain carotenoids in nonphotosynthetic tissues, such as flowers and fruits. Here, their function has been attributed to making these tissues more attractive to animals, thereby promoting pollination and seed dispersal. Carotenoids are sometimes found in the endosperm of monocotyledonous grains such as maize (MANGELSDORF and FRAPS 1931). These carotenoids function in part as precursors to the phytohormone abscisic acid (ABA) (PARRY and HORGAN 1991). Mutant lines of maize that accumulate carotenoid precursors and do not produce ABA fail to become dormant and germinate on the cob (reviewed by ROBERTSON 1975). This precocious germination is referred to as vivipary.

The first C_{40} carotenoid is produced by the tail-to-tail condensation of two molecules of geranylgeranyl pyrophosphate (GGPP) to form the colorless compound phytoene. GGPP is also a precursor to phytohormones and the phytyl side chain of phylloquinones, tocopherols and chlorophylls (GOODWIN 1971). In addition, isoprenylations of proteins often involve geranylgeranyl moieties (RANDALL *et al.* 1993). The regulation of the flow of GGPP into the biosynthetic pathways that lead to these varied and essential molecules is a key event in cellular metabolism. The enzyme that commits GGPP to the carotenoid biosynthetic pathway is phytoene synthase. This enzyme has been isolated from red pepper (*Capsicum annuum*; DOGBO *et al.* 1988). Phytoene synthase genes or cDNAs have been isolated from red pepper (ROMER *et al.* 1993), *Arabidopsis thaliana* (SCOLNIK and BARTLEY 1994), *Cucumis melo* (KARVOUNI *et al.* 1995), *Neurospora crassa* (SCHMIDHAUSER *et al.* 1994), and several prokaryotes (ARMSTRONG *et al.* 1990; MISAWA *et al.* 1990; CHAMOVITZ *et al.* 1992). In tomato, two different genes for phytoene synthase, *psy1* and *psy2*, have been isolated (BARTLEY *et al.* 1992; BARTLEY and SCOLNIK 1993). Both *psy1* and *psy2* were found to be expressed in seedlings, mature leaves and fruits of tomato. *psy1* transcripts predominate in seedlings and fruits, while *psy2* is the more active gene in mature leaves (BARTLEY and SCOLNIK 1993).

The only carotenoid biosynthetic gene isolated from a monocotyledonous plant is the *y1* gene of maize (BUCKNER *et al.* 1990). At least three types of alleles of *y1* are known: dominant alleles of *y1*, which condition a yellow endosperm/green leaf phenotype, recessive alleles of *y1*, which condition a white to pale-yellow endosperm/green leaf phenotype, and pastel alleles of *y1*, which condition white to pale-yellow endosperm/pale-green leaf phenotypes. The yellow endosperm is due to an accumulation of carotenoids; the pale-green leaf phenotype of the pastel alleles is a consequence of a reduction of carotenoid levels in the leaf resulting in the photooxidation of the chlorophyll in the seedling. The *y1* pastel alleles are temperature sensitive (*i.e.*, elevated temperatures are necessary for, or exaggerate, this phenotype). The phenotypes associated with these

Corresponding author: Brent Buckner, Division of Science, Northeast Missouri State University, Kirksville, MO 63501.
E-mail: sc12%nemomus@academic.nemostate.edu

three types of alleles suggest that the *y1* gene is involved in carotenoid biosynthesis in both the leaf and endosperm of maize. In this study, we report sequences of genomic and cDNA clones of the *y1* gene of maize. From these data, we have deduced that the *y1* gene product is phytoene synthase. We also describe insertion polymorphisms that flank the locus and the transcript/expression patterns of this gene in *Y1* and *y1* tissues.

MATERIALS AND METHODS

Genetic material: The maize line that contained a dominant allele of *y1*, and was used in carotenoid and *y1* mRNA analyses, was a hybrid of inbred lines Q66 and Q67 (designated Q60). The maize lines employed in this study that contain a recessive allele of *y1* or the pastel allele *y1-2053* were previously described (BUCKNER *et al.* 1990).

Extraction and analysis of carotenoids: Ears were harvested within a 2-day period 30 days after pollination (DAP). Kernels were removed and stored under argon at −20°. At the time of analysis, embryos were removed from kernels and extracted separately from the remainder of the kernel (designated "endosperm"). Approximately 20 embryos or 10 endosperms were pooled and carotenoids were extracted using the method of BARUA *et al.* (1989). Samples from three ears of both genotypes (*Y1Y1Y1/Y1Y1* or *y1y1y1/y1y1* endosperm/embryo) were analyzed in triplicate with the exception of the *Y1Y1* embryos, which were analyzed twice.

Carotenoids were separated by isocratic reversed-phase HPLC as described by BARUA *et al.* (1993). The Waters high-performance liquid chromatography (HPLC) system (Milford, MA) used included a model 501 pump, a model 484 detector set at 450 nm, a U6K injector, a system interface module and Baseline Software (Dynamic Solutions, Ventura, CA) for integration and analysis of chromatograms. Since this system did not completely separate lutein and zeaxanthin, the values for these two carotenoids are reported together. Carotenoids were identified by their elution times relative to standards and by their absorption spectrum. The amount of each carotenoid present in embryo and endosperm samples was calculated from a standard curve produced for α-carotene, β-carotene, β-cryptoxanthin and lutein. The extinction coefficients used for this process were obtained from BARUA and FURR (1993). For use as standards, α-carotene and β-cryptoxanthin were extracted and purified from carrot and dried peaches, respectively, whereas β-carotene (General Nutrition Centers, Pittsburgh, PA) and lutein (Sigma Chemical Company, St. Louis, MO) were further purified from commercial sources.

RNA blot hybridization analysis: Total RNA was isolated from 1-wk-old seedling leaves or 30 DAP embryo or endosperm by the method of MCCARTY (1986) or by using an Extract-A-Plant RNA Isolation Kit (Clontech Laboratories, Inc., Palo Alto, CA). PolyA RNA was purified by using a PolyATtract mRNA Isolation System (Promega Corporation, Madison, WI). Approximately 10 μg of polyA RNA was electrophoretically separated through a 1.0% agarose slab gel, blotted to nylon membranes and hybridized to ^{32}P labeled DNA as described by THOMAS (1980). The hybridization probe was the 3.5-kb *Bam*HI fragment of *y1* that has been previously described (BUCKNER *et al.* 1990).

Complementary DNA cloning: A maize cDNA library prepared from inbred B73 plants at the five leaf stage (Stratagene Inc., La Jolla, CA) was screened by plaque hybridization (MASON and WILLIAMS 1987) using the 3.5-kb *Bam*HI fragment of *y1*. A positive plaque was purified, the DNA from this bacteriophage was isolated (DAVIS *et al.* 1986), and the 1151-bp maize insert was subcloned into pGem7zf(+) (Promega Corporation, Madison, WI) and designated pl12A33A. A second *y1* cDNA clone, plcDNA1.1, which contained a 1366-bp maize insert, was isolated using the same protocols from a cDNA library that was prepared from 30 DAP whole maize kernels of color-converted inbred line W22 (provided by Dr. KAREN C. CONE, University of Missouri-Columbia).

DNA sequencing: The 3.5- and 2.4-kb *Bam*HI subclones of *y1* previously described (BUCKNER *et al.* 1990) and plcDNA1.1 were completely sequenced on both strands. This sequencing was facilitated by processing large subclones into sets of nested-deletion subclones by restriction enzyme digests or by exonuclease III deletion. The ends of the resulting linear DNA fragments were made blunt using T4 DNA polymerase or S1 nuclease and Klenow DNA polymerase, self-ligated in the presence of 5% PEG, and transformed into competent DH5α *Escherichia coli*. Double stranded plasmid DNAs were prepared and sequenced using appropriate vector primers for dideoxy sequencing (SANGER *et al.* 1977). Synthetic oligonucleotides were prepared and used to complete the sequences of both strands where the available deletion subclones proved inadequate for this task. A subclone that contained the junction between the 3.5- and 2.4-kb *Bam*HI genomic *y1* subclones was partially sequenced to determine that the genomic fragments were contiguous. The cDNA clone pl12A33A was also partially sequenced, to obtain additional 5′ and 3′ mRNA sequence not represented in plcDNA1.1. Sequencing reactions and gels were as previously described (JIN and BENNETZEN 1994). Sequences were analyzed using the GCG software programs (DEVEREUX *et al.* 1984) run on Purdue's VAX cluster via the AIDS Center Laboratory for Computational Biochemistry.

Analysis of mRNA ends: Rapid amplification of cDNA ends (RACE) reactions were performed as described (FROHMAN *et al.* 1988; LOH *et al.* 1989; JIN and BENNETZEN 1994) on 30 DAP *Y1Y1Y1* endosperm mRNA. The 34 base oligo(dT) primer used for cDNA synthesis in 3′ RACE was (5′-GACTCGAGTCGACATCGTTTTTTTTTTTTTTTT-3′). The first polymerase chain reaction (PCR) amplification of the cDNA was performed with a *y1*-specific primer from exon 5 (5′-AGAGGCAGATCAAGAGGGCC-3′) and an adaptor primer (5′-GACTCGAGTCGACATCG-3′). This reaction was electrophoretically resolved on a 1.5% agarose gel, and cores were taken at 3-mm intervals through the length of the lane with a micropipette tip. The cores were each eluted into 200 μl of sterile deionized water and 1–5 μl was reamplified with an overlapping *y1*-specific primer (5′-GGCCAGGATGTTTTTTGAG-3′) and the same adaptor primer. The reactions were electrophoretically resolved on a 1.5% agarose gel containing ethidium bromide. The gel was illuminated with a long wave (365 nm) UV lamp, and individual bands were excised with a razor blade and glass powder purified. The purified DNA fragments were digested with *Sst*I, which cuts in exon 5, and *Cla*I, which cuts the adaptor primer, agarose gel purified and directionally cloned into pBluescriptSK (Stratagene, Inc.). Clones were sequenced to determine polyadenylation sites.

5′ RACE was performed on the same mRNA, but cDNA was prepared using random hexamers and tailed with deoxyguanosine using deoxynucleotidyl transferase (Pharmacia Biotech Inc., Piscataway, NJ). An oligo(dC) primer (5′-GGAATTCGCGGCCGCCCCCCCCCCCCCCCCA/G/T-3′) and a *y1*-specific primer from exon 3 (5′-CAAGCATGTCGTAAGGACG) were used to prime a first PCR reaction. The same oligo(dC) primer and either oligo p23007 (5′-GCTTGAGCACGACGTCCTAG-3′) or oligo 41E (5′-CGTCTTGAGCAGGGTGGAG-3′) were used to reamplify an aliquot of the first

TABLE 1

Distribution of carotenoids in maize kernels

Tissue	Carotenoids		
	Lutein/zeaxanthin	β-Cryptoxanthin	β-Carotene
Y1Y1 Embryo[a]	16.54 ± 3.20	3.86 ± 1.14	2.07 ± 0.43
y1y1 Embryo[a]	12.65 ± 1.14	2.80 ± 0.80	2.72 ± 1.03
Y1Y1Y1 Endosperm[a]	10.00 ± 0.59	0.54 ± 0.07	2.02 ± 0.07
y1y1y1 Endosperm[a]	0.62 ± 0.32	0.23 ± 0.03	0.29 ± 0.06

Carotenoids were extracted from maize tissues and analyzed by HPLC as described under MATERIALS AND METHODS. Each value represents the mean ± SD of embryo or endosperm from kernels of three separate ears measured in triplicate, except the *Y1Y1* endosperms, which were done in duplicate. Data are expressed as *micrograms* per gram of wet tissue.

[a] Values were compared by using an unpaired *t*-test. All differences between *y1* and *Y1* embryos were not significant, while the differences between *y1* and *Y1* endosperms were significant ($P < 0.001$).

reaction. This amplification reaction was resolved through a 1.5% agarose gel and filter hybridized to the 2.4-kb *Bam*HI fragment of the genomic *y1* clone. Bands that hybridized were isolated and directly sequenced.

RESULTS

Distribution of carotenoids in the maize kernel: To analyze the distribution of carotenoids in the maize kernel, we extracted total carotenoids separately from endosperms and embryos of kernels that were homozygous for either a dominant (yellow kernel) or recessive (white kernel) allele of *y1*. The carotenoids were separated and quantified by HPLC. The major carotenoids found in the kernels were zeaxanthin/lutein, a xanthophyll monoester, β-cryptoxanthin, α-carotene and β-carotene. The types and amounts of carotenoids were similar in the embryos of kernels that were homozygous for either the dominant or recessive alleles of *y1* (Table 1). The types of carotenoids present in the endosperm of kernels homozygous for the dominant or recessive alleles of *y1* were also the same; however, they were found in significantly lower quantities in the endosperm of homozygous recessive kernels (Table 1).

Expression of the *y1* gene: RNA blot hybridization analyses were performed to investigate the expression of the *y1* gene. A 1.8-kb mRNA that hybridized to a *y1* probe was observed in the leaves of 1-wk-old seedlings which were homozygous for a dominant allele of *y1* (Figure 1). The leaves of 1-wk-old seedlings that were homozygous for a pastel allele of *y1*, *y1-2053*, contained a 1.7-kb *y1*-homologous mRNA. As previously shown (BUCKNER *et al.* 1990), 1-wk-old seedlings that were homozygous for the recessive allele of *y1* contained a 2.0-kb mRNA with homology to *y1*.

Embryos isolated from kernels 30 days after pollination (DAP) also exhibited *y1* mRNA in all genotypes examined. The allele-specific length of the *y1* mRNA detected in the embryo is consistent with that found in 1-wk-old seedlings leaves (Figure 1).

A 1.8-kb *y1*-homologous transcript was also detected in endosperm isolated from 30 DAP kernels which were homozygous for a dominant allele of *y1*. However, no *y1*-hybridizing transcript was detected in 30 DAP endosperm that was homozygous for the recessive allele of *y1* (Figure 1). Rehybridization of the RNA blots used in this study with a human actin probe (SOLOMON and RUBENSTEIN 1987) indicated that there were similar amounts of RNA in both endosperm samples (Figure 1). Densitometric analysis of these autoradiographs indicated that the transcript produced from the dominant *y1* allele was approximately five times more abundant in leaves than in endosperm.

Sequences of y1 genomic and cDNA clones: The

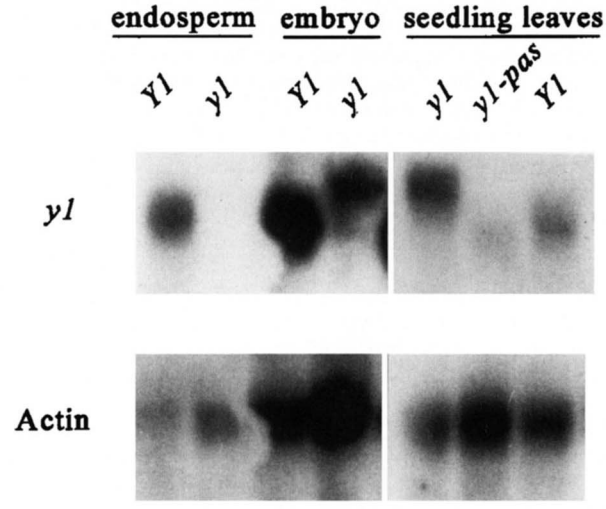
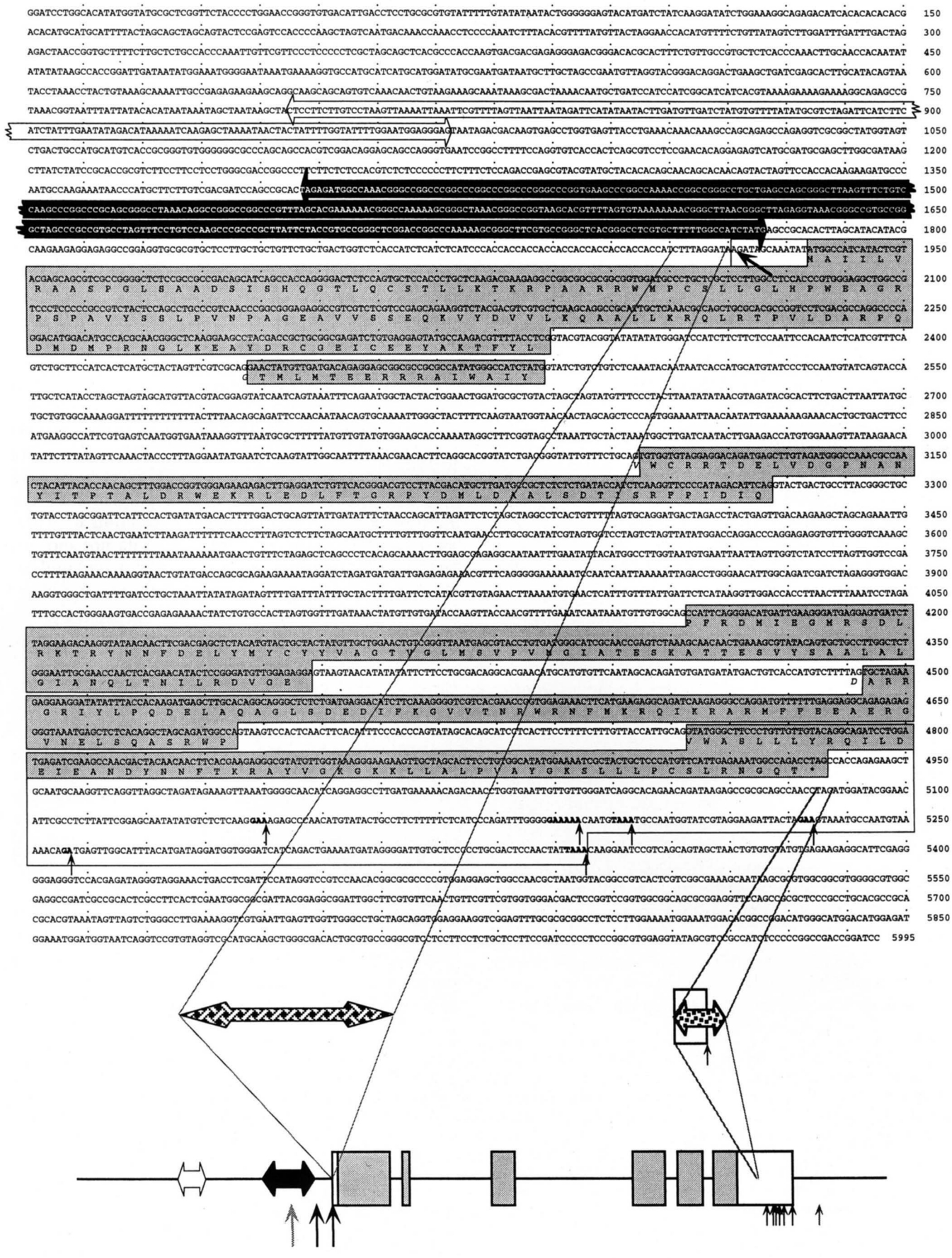


FIGURE 1.—*y1* RNA accumulation in different tissues and different *y1*-allelic backgrounds. Autoradiographs of RNA blot hybridization analyses. PolyA RNA was isolated from 30 DAP endosperm and embryo, or 1-wk-old seedling leaves that were homozygous for the designated allele. Approximately 10 μg were loaded per lane and subjected to electrophoresis. Northern blots were prepared and hybridized with radiolabeled *y1* and actin DNA. Sizes of hybridizing RNAs are as indicated in the text. The *y1-2053* allele is designated *y1-pas*. Due to the lower relative abundance of *y1* transcript in endosperm and embryo as compared with leaf, a longer exposure of the RNA blot hybridization is shown for these samples.



cloning of the *y1* gene was previously described (BUCKNER *et al.* 1990). We now have cloned 1151- and 1366-bp cDNA homologues of *y1* transcripts isolated from leaves of line B73 and 30 DAP kernels of line W22, respectively. We have sequenced all of the 1151-bp cDNA, the ends of the 1366-bp cDNA, and all of a contiguous 5995-bp that contains the genomic *y1* gene (Figure 2). The cDNA sequences have good homology to the tomato cDNAs for phytoene synthase (*psy*). In addition, the deduced amino acid sequence of the protein encoded by the *y1* gene also has good homology to the plant PSY proteins: 78% similarity, 67% identity with tomato PSY1; 81% similarity, 69% identity with tomato PSY2; 80% similarity, 71% identity with the red pepper PSY protein; and 80% similarity, 68% identity with an Arabidopsis PSY (Figure 3). We conclude from this strong homology that the maize *y1* gene codes for phytoene synthase.

The *y1* gene was isolated by using a transposon tagging and cloning strategy (BUCKNER *et al.* 1990). The mutant allele that was initially cloned (*y1-2053*) was induced by a *Mu3* element of the Robertson's *Mutator* family of transposons. Sequencing of the *y1* DNA flanking one side of the *Mu3* element in *y1-2053* indicated that the element is inserted at position 1911 in our genomic sequence. The apparent target site for this insertion was CTTTAGGAT. This sequence has homology at TTTA and AT to the GATTTATAT sequence, which was duplicated in the only other *Mu3* insertion so far characterized, one in which *Mu3* inserted into the TATA box of the *adh1* gene (CHEN *et al.* 1987). Moreover, this sequence has reasonable homology (4/9) to the target consensus sequence identified by CRESSE and coworkers (CRESSE *et al.* 1994) for the insertions of *Mu1*-related elements.

By comparing the *y1* cDNA and genomic sequences, we have identified the putative exons and introns of the maize *y1* gene (Figure 2). The observed intron positions are identical to those for the tomato *psy1* gene (RAY *et al.* 1987). A full-length cDNA from the dominant allele of *y1* would be expected to be 1.8 kb in length. Hence, the longest (1366-bp) *y1* cDNA that we have analyzed does not extend to the 5′ end of the transcript. By comparison to the coding region of the tomato *psy1* gene (RAY *et al.* 1987), the longest maize cDNA that we have sequenced is 5′-truncated within the first exon.

Mapping the 5′ and 3′ ends of *y1* transcripts: The RACE technique was used to accurately position the 5′ and 3′ ends of *y1* mRNAs isolated from the hybrid Q60. PolyA RNA from 30 DAP *Y1Y1Y1* endosperms yielded at least six *y1* 3′ ends, with mapped polyadenylation sites at nucleotides 5138, 5189, 5200, 5232, 5256 and 5342 in our sequence (Figure 2). At least one longer 3′ end was also observed, but was not sequenced. Hence, the choice of a polyadenylation site is variable at *y1* in 30 DAP endosperm of these *Y1Y1Y1* kernels. The various sequenced ends all map a few hundred base pairs distal to the TAG sequence at 4934 that terminates the last long *y1* open reading frame of *y1* (Figure 2).

The 5′ RACE reactions yielded a number of products, the shortest of which terminated at nucleotide 1921 in our genomic sequence (Figure 2). This shortest molecule represents a true mRNA 5′ end, since the capping G nucleotide was copied into the cDNA (data not shown). At least two longer 5′ RACE products were also observed to terminate within 300-bp upstream of position 1921; one approximately at position 1825 and another ~200-bp further upstream (data not shown). Hence, *y1* transcription has multiple start sites and/or variable 5′ end processing in 30 DAP endosperm of these *Y1Y1Y1* kernels.

The first ATG downstream from position 1921, where our shortest *y1* transcript initiates, is at position 1934 of our genomic sequence. This ATG is in frame with the phytoene synthase-homologous portion of the gene, and no other ATG is found within 150 bp upstream. Hence, we predict that the ATG at 1934 encodes the translation start site for the *y1* gene product. This would predict a *y1*-encoded polypeptide of 46.5 kD. This molecular weight is in close agreement with the estimated molecular weights of the tomato (estimated at 47 and 46.7 kD; BARTLEY *et al.* 1992 and RAY *et al.* 1987, respectively) and *Capsicum* (48 kD; DOGBO *et al.* 1988) PSY proteins.

Elements in the *y1* promoter and 3′ end regions: Two insertion sequences were found just 5′ to the protein-encoding portion of *y1* (Figure 2). The first insertion sequence, *Ins2*, is delineated by the imperfect inverted repeats TAGAGATGGCCAAA and ATAGATGGCCAAA.

FIGURE 2.—Sequence and map of the *y1* locus. The primary sequence of the *y1* locus cloned from hybrid Q60 is shown, with various landmarks highlighted. The open arrow covering nucleotides 795–972 is a *Stowaway* element. The dark arrow covering nucleotides 1397–1775 is an *Ins2* element. The boxes show exons, starting from the RNA with the shortest leader sequence mapped and going to the longest 3′ end mapped; unshaded regions are the leader and trailer while the shaded portions of the box are predicted peptide-encoding regions of the exons. Letters below the DNA sequence are the indicated amino acid sequence. The slanted arrow at 1921 denotes the precisely mapped 5′ end of the shortest *y1* transcript identified by 5′ RACE. The smaller arrows indicate several of the different 3′ ends of *y1* transcripts identified by 3′ RACE; the bold nucleotides are meant to show that the A base or bases in the DNA sequence at these sites make it impossible to determine which encoded nucleotide actually was at the 3′ end before polyA addition. Below the sequence portion of the figure is a schematic showing these same results, plus additional 5′ and 3′ transcript ends that were not precisely mapped. The crosshatched arrow above the schematic indicates the size and position of the *Mu3* insertion in *y1*-2053. The block-filled arrow above the schematic denotes the *Tourist* element insertion in the 3′ end trailer, and a polyA addition site, in the *y1* allele from maize line B73. The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession number U32636.

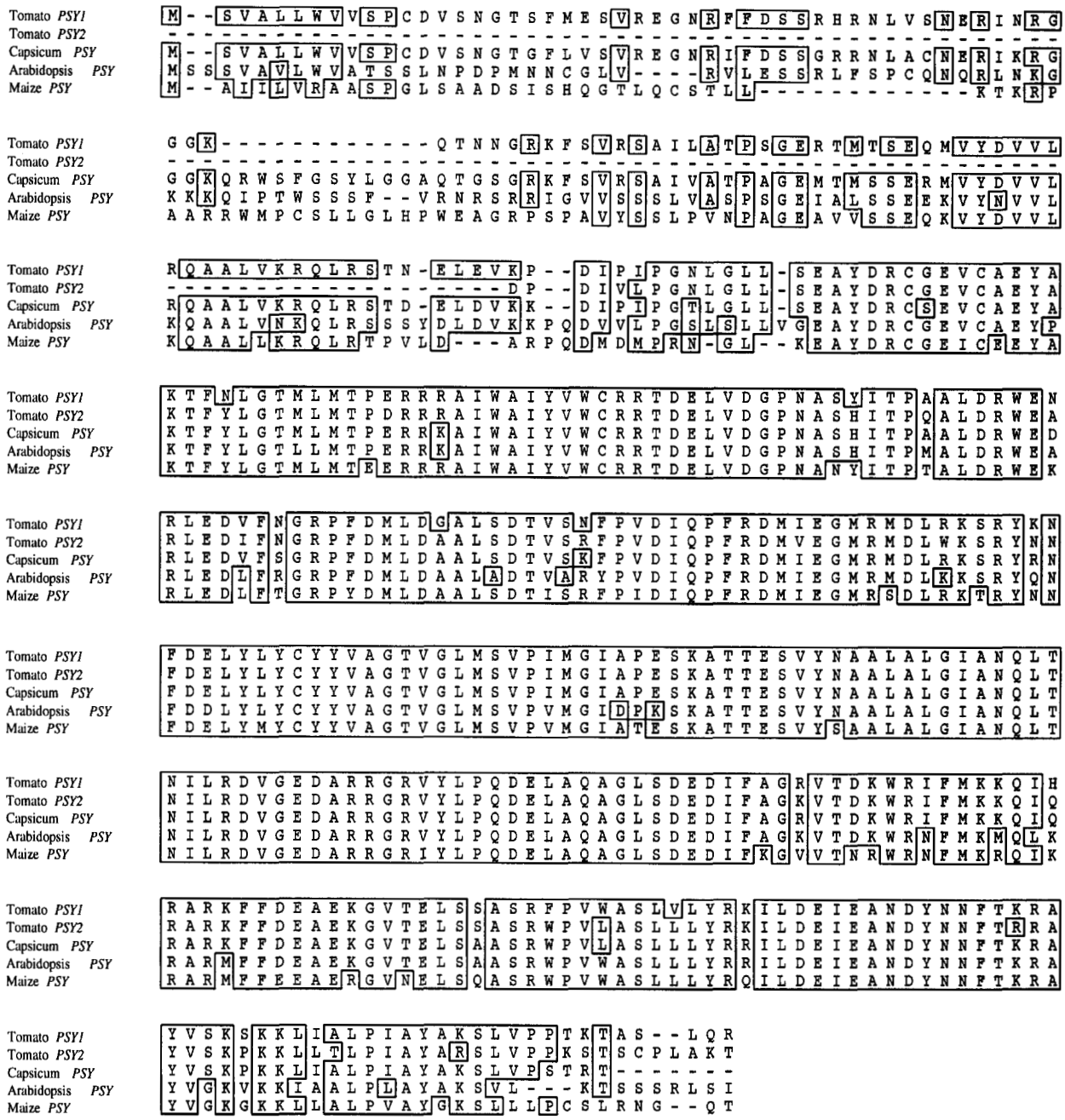


FIGURE 3.—Comparison of predicted plant phytoene synthase (PSY) sequences. Boxed regions indicate high sequence similarity, as defined by the default setting on the Pileup program for the GCG sequence analysis system (DEVEREUX *et al.* 1984). The sequences of the clones that yielded predicted proteins for tomato, Capsicum and Arabidopsis PSYs have been published (BARTLEY *et al.* 1992; BARTLEY and SCOLNIK 1993; ROMER *et al.* 1993; SCOLNIK and BARTLEY 1994), and the maize PSY sequence is from the DNA sequence presented in Figure 2.

Insertion of this 379-bp sequence has created an 8-bp direct repeat, AGCCGCAC, flanking the element. *Ins2* has been identified 5′ to a *bz1* allele of maize (RALSTON *et al.* 1988), and is also found in a region 3′ to a *c1* allele of maize (PAZ-ARES *et al.* 1987). Some evidence indicates that the transposition of *Ins2* may be regulated by the *Ac* transposable element (D. SCHWARTZ, unpublished observation). However, in *Ac*-containing lines, our PCR analysis did not uncover any empty donor sites expected with excision of the *Ins2* element within *y1*. This study also identified dominant *y1* alleles (which condition a yellow endosperm) that do not contain an *Ins2* insertion at this site (P. SAN MIGUEL and J. L. BENNETZEN, unpublished results).

The second insertion sequence in the *y1* promoter region is a member of the *Stowaway* transposable element family (BUREAU and WESSLER 1994a). This sequence is bounded by the imperfect inverted repeats CTCCTTC and CTCCCTC, and created a flanking direct duplication of 2 bp, TA.

At positions 1877–1909 in our genomic sequence, the trinucleotide CCA is repeated 11 times. Trinucleotide repeats are common in plants, as are other simple sequence repeats (MORGANTE and OLIVIERI 1993). When the EMBL and GenBank data bases were screened for the presence of the 10 different trinucleotide repeats, the CCA repeat was found to represent 5% of the repeats detected. In all higher eukaryotes examined, these simple sequence repeats are both present and highly variable between lines. In PCR analysis of other dominant (yellow endosperm) *y1* alleles, we have found that this repeat exhibits allele-specific repeat number variation, as indicated by the sizes and sequences of the amplification products (PHELPS and BUCKNER 1995).

The cDNAs that we have sequenced were derived from maize lines B73 and W22. These inbred lines are homozygous for dominant alleles of *y1*. The genomic clone that we sequenced is a dominant allele of *y1* isolated from the maize stock Q60, a hybrid of inbred lines Q66 and Q67. During our sequence comparisons, we found only a single base pair difference between these *y1* alleles in the coding regions, a C (B73) to A (Q60) transversion at position 4657 (Figure 2). However, we have found that the 194 bp preceding the polyadenylated tail of the B73 cDNA is not found within the sequenced genomic DNA. We have analyzed the genomic *y1* DNA sequence for ~3 kb downstream from this loss of homology (Figure 2) and have not found this sequence. Restriction mapping to compare B73 and Q60 *y1* alleles in this region identified an ~300-bp insertion in the *Y1+B73* allele (data not shown and Figure 2). Using oligonucleotide primers flanking this polymorphism, the *y1* 3′ genomic sequence of B73 was PCR amplified and sequenced (data not shown). The results indicated a 343-bp insertion at position 5085/5087 in our genomic sequence, which is flanked by a 3-bp direct repeat of target DNA, TAG (Figure 2). This insertion element has imperfect inverted repeats, and qualifies as a member of the D subfamily of *Tourist* transposons (BUREAU and WESSLER 1994b). This element should not contribute to the protein-encoding potential of the *Y1+B73* allele, but does provide the polyadenylation site that was identified in the *Y1+B73* cDNA, pl12A33A.

DISCUSSION

y1 encodes phytoene synthase: The *y1* gene of maize has long been known to be involved in the biosynthesis of carotenoids (MANGELSDORF and FRAPS 1931). Although this gene had been previously cloned (BUCKNER *et al.* 1990), its function was not determined. The sequence of *y1* genomic and cDNA clones allowed us to deduce the amino acid sequence of the protein encoded by *y1*. This protein has significant amino acid identity to plant phytoene synthase (Figure 3), the enzyme that catalyzes the first biosynthetic step specific to carotenoid production. We conclude that the maize *y1* gene codes for phytoene synthase.

DNA, RNA and functional polymorphisms at *y1*: Several interesting alleles of *y1*, with different phenotypes, have been isolated by maize scientists during a period of many years. Our analyses of *y1* expression at the level of RNA accumulation indicates that these alleles also differ in the level, tissue-specificity and sizes of transcripts synthesized. For example, across all tissue types examined, the dominant *y1* allele analyzed in this study produced a 1.8-kb mRNA, a *y1* pastel allele (*y1-2053*) produced a 1.7-kb mRNA and a standard recessive *y1* allele produced a 2.0-kb mRNA. The mRNA length differences found between the dominant, pastel and recessive alleles of *y1* are apparently a result of allelic differences in transcript initiation and/or termination sites. In addition, even in a single tissue (*e.g.*, endosperm), different 5′ and 3′ ends were found for the *y1* transcripts.

Since none of these alleles have a defined parent/progeny relationship, it is not clear how many of these differences are directly due to functional mutations at *y1* and how many are correlated but functionally insignificant polymorphisms. For instance, the sequence of the *y1* genomic clone revealed several interesting classes of repeated DNA (*Ins2*, *Stowaway*, and *Tourist* elements, plus 11 tandem copies of the trinucleotide CCA) inserted either 5′ or 3′ to the structural gene. The 3′ *Tourist* element provides a polyA addition site in some *y1* alleles, but is not present in others. At the 5′ end of the gene, an *Ins2* element provides an alternative transcription start site and expected promoter features for some dominant *y1* alleles, yet is missing from other dominant *y1* alleles that also provide apparent "wild-type" function. Similarly, variable numbers of the CCA repeats that are found within the 5′ leader in some transcripts, and just before the 5′ end of others, do not detectably affect wild-type function of the locus (PHELPS and BUCKNER 1995). Therefore, various allele-specific insertions, and/or trinucleotide repeat number variability may account for some of the RNA length polymorphism observed. These data suggest a tremendous amount of plasticity to promoter function at *y1* and may also provide some insight into why transposable elements are not more detrimental to plant fitness.

***y1* expression:** Studies of carotenoid composition, levels, and distribution in various *y1* allelic backgrounds are also limited by the lack of a direct parent/progeny relationship between wild-type and mutant alleles. However, the correlations between *y1* RNA synthesis, carotenoid production, and phenotype are informative. Plants that possess a dominant allele of *y1* contain carotenoids in their photosynthetic tissue, embryo and endosperm. Plants homozygous for a standard recessive allele of *y1* still produce relatively normal levels of carotenoid pigments and *y1* mRNAs in the seedling and embryo, but produce endosperm with greatly reduced

amounts of carotenoids in comparison with plants that have a dominant allele of *y1*. This tissue-specific effect of the recessive *y1* mutation on carotenoid production is consistent with the white endosperm phenotype and with our inability to detect *y1* mRNA by RNA blot hybridization analysis in *y1y1y1* endosperm.

y1 mRNAs are most abundant in seedlings, where carotenoids are essential for protection of chlorophyll from photooxidation. It is not surprising that the embryo of both white and yellow kerneled maize contain carotenoid pigments and *y1* mRNA, as the carotenoids are precursors to a hormone (ABA) that is required for maintenance of embryo dormancy. However, the biological significance of carotenoid accumulation in the endosperm of maize is not known.

Despite the absence of detected *y1* mRNA in 30 DAP endosperm of plants that were homozygous for the standard recessive alleles of *y1*, low levels of carotenoids were detected in this tissue. This suggests that a second phytoene synthase gene may be active in this tissue and perhaps in others as well. Alternatively, the inability to detect *y1* mRNA in endosperm that is homozygous for recessive *y1* may be explained if the gene is expressed in this tissue at an earlier time or at a low level below the sensitivity afforded by RNA blot hybridization. Using RNA blot hybridization analysis, *psy1* expression could only be detected in the ripened fruit of tomato (RAY *et al.* 1992), even though carotenoids were clearly present in other photosynthetic tissues. Using a reverse transcriptase and polymerase chain reaction assay for expression of the *psy1* gene, GIULIANO *et al.* (1993) were able to detect *psy1* transcripts in all tomato tissues investigated. The tomato *psy1* gene was expressed at levels ~35 and 75 times as high in the petals and fruit, respectively, compared with seedlings. *psy1* transcripts also were detected at very low levels (*i.e.*, one-tenth of seedling levels) in the roots of tomato plants. Similar studies in maize should indicate if and when the recessive allele of *y1* is transcribed in the endosperm of *y1y1* plants.

Mutational lesions at *y1*: The pastel allele *y1-2053* was caused by an insertion of a *Mu3* element into one of the dominant *y1* alleles of hybrid Q60. The relatively short (1.7 kb) *y1* mRNA detected by RNA blot hybridization in *y1*-2053 plants could be a consequence of the *Mu3* element causing the choice of alternative sites for transcriptional initiation site or mRNA splicing. Alterations in RNA processing and acquisition of new promoters sited within the element have both been observed in other *Mutator* insertion mutations (STROMMER and ORTIZ 1989; LUEHRSEN and WALBOT 1990; BARKAN and MARTIENSSEN 1991). This shorter *y1* transcript may not be efficiently translated or it may be translated into a less functional PSY protein resulting in the reduced endosperm carotenoids and temperature-sensitive seedling color associated with this pastel allele of *y1*.

The pastel mutant phenotype is by far the most common type of *y1* mutation. ROBERTSON (1985) generated several hundred *Mutator*-induced mutations at *y1*, of which ~75% were pastel alleles (D. S. ROBERTSON, personal communication). In addition, pastel alleles of *y1* have been reported that appear to be generated by gamma irradiation (ROBERTSON and ANDERSON 1961), while others were found as spontaneous mutations (D. S. ROBERTSON, personal communication). *Mutator*-induced pastel alleles of *y1* are known in which the genetic lesion is not in the same region of the gene as the *y1-2053* allele described in this study (BUCKNER *et al.* 1990). Therefore, generalizations about the location of the genetic lesion responsible for the pastel phenotype can not be made from the data regarding *y1-2053*.

The recessive *y1* allele examined in this study appears to be a regulatory mutation, permitting relatively normal levels of *y1* mRNA accumulation in the seedling and embryo (where it is essential), while conditioning undetectable levels in the endosperm. Hence, it is likely to be due to a deficiency in either RNA synthesis or stability. It is not known if this relatively normal expression in the seedling and deficiency in the endosperm at the RNA level is true of most or all standard recessive *y1* alleles.

Mutations that fully inactivate *y1* function in all tissues may be missing from the extensive *y1* mutant collection due to a presumed lethality of this phenotype. The absence of carotenoids from the leaf would lead to an albino/viviparous phenotype that is a seedling lethal. These mutations are common in maize and have been mapped (ROBERTSON 1975), but none are reported to be allelic to *y1*. It is also possible that a complete inactivation of *y1* would have earlier lethal effects and, therefore, null alleles of *y1* might be difficult to identify. BUCKNER and REEVES (1994) have demonstrated that female gametophytes that are deficient for the *y1* locus are viable but exhibit reduced fitness. Therefore, it is possible that *y1* may perform a vital role in early embryo development.

Phytoene synthase gene number in maize: One possible interpretation of the very low levels of endosperm carotenoids in recessive *y1* mutations that lack detected *y1* endosperm RNA is that additional phytoene synthase genes are present in the maize genome. DNA gel blot hybridization does detect additional weak bands of cross-hybridization that could indicate a *y1* homologue or homologues (BUCKNER *et al.* 1990). Moreover, the presence of a single copy per haploid genome of a gene for an essential function is unusual in higher eukaryotes and is particularly unexpected for an organism like maize that is derived from an ancient polyploidization (HELENTJARIS *et al.* 1988). Tomato does have more than one expressed phytoene synthase gene (BARTLEY *et al.* 1992; BARTLEY and SCOLNIK 1993). For all these reasons, we expect that maize will have other phytoene synthase genes beyond *y1*. However, the combined physiological, molecular and genetic results generated for the *y1* alleles indicate that *y1* provides the primary, perhaps

only, phytoene synthase activity in the endosperm and may be equally important in the leaf.

Postulating additional phytoene synthase loci could explain why no albino or viviparous alleles of *y1* have been described. In this scenario, another phytoene synthase gene or genes would be active in the leaf and embryo to prevent chlorophyll photooxidation and to allow the hormone synthesis needed to establish seed dormancy, respectively. Alternatively, a *y1* mutation that fully loses function in all tissues might be an embryonic lethal. This latter model is supported by the observation that some *y1*-pastel mutations show equivalent degrees of pigmentation deficiency in both the seedling and endosperm. If these pastel mutations are due to a decrease in phytoene synthase activity, this suggests that no other gene in the seedling leaf can provide enough enzyme to overcome this deficiency or that the altered *y1*-pastel enzyme forms an inhibitory complex with other phytoene synthases. The wealth of alleles at the *y1* locus, and the availability of cloned *y1* sequences as probes, should now allow more detailed characterization of the relationships between phytoene synthase gene activity, carotenoids, and phenotypes in a variety of maize tissues and backgrounds.

We thank Dr. K. C. CONE for providing us with a cDNA library and A. BALLMANN and A. ANNUAR for technical assistance. This research was supported by grants from The Rockefeller Foundation's International Program on Rice Biotechnology (No. 90031, #123), and the National Institutes of Health (No. A127713) in support of Purdue's AIDS Center Laboratory for Computational Biochemistry.

LITERATURE CITED

ARMSTRONG, G. A., M. ALBERTI and J. E. HEARST, 1990 Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. Proc. Natl. Acad. USA **87:** 9975–9979.

BARKAN, A., and R. A. MARTIENSSEN, 1991 Inactivation of maize transposon *Mu* suppresses a mutant phenotype by activating an outward-reading promoter near the end of *Mu1*. Proc. Natl. Acad. Sci. **88:** 3502–3506.

BARTLEY, G. E., and P. A. SCOLNIK, 1993 cDNA cloning, expression during development, and genome mapping of *PSY2*, a second tomato gene encoding phytoene synthase. J. Biol. Chem. **268:** 25718–25721.

BARTLEY, G. E., P. V. VIITANEN, K. O. BACOT and P. A. SCOLNIK, 1992 A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthetic pathway. J. Biol. Chem. **267:** 5036–5039.

BARUA, A. B., and H. C. FURR, 1993 Extraction and analysis by high-performance liquid chromatography of carotenoids in human serum. Methods Enzymol. **213:** 273–281.

BARUA, A. B., R. O. BATRES, H. C. FURR and J. A. OLSON, 1989 Analysis of Carotenoids in human serum. J. Micronutr. Anal. **5:** 291–302.

BARUA, A. B., H. C. FURR, D. JANICK-BUCKNER and J. A. OLSON, 1993 Simultaneous analysis of individual carotenoids, retinol, retinyl esters, and tocopherols in serum by isocratic non-aqueous reversed-phase HPLC. Food Chem. **46:** 419–424.

BUCKNER, B., and S. L. REEVES, 1994 Viability of female gametophytes that possess deficiencies for the region of chromosome 6 containing the *Y1* gene. Maydica **39:** 247–254.

BUCKNER, B., T. L. KELSON and D. S. ROBERTSON, 1990 Cloning of the *y1* locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell **2:** 867–876.

BUREAU, T. E., and S. R. WESSLER, 1994a *Stowaway*: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell **6:** 907–916.

BUREAU, T. E., and S. R. WESSLER, 1994b Mobile inverted repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. Proc. Natl. Acad. Sci. USA **91:** 1411–1415.

CHAMOVITZ, D., N. MISAWA, G. SANDMANN and J. HIRSCHBERG, 1992 Molecular cloning and expression in *Escherichia coli* of a cyanobacterial gene coding for phytoene synthase, a carotenoid biosynthetic enzyme. FEBS Lett. **296:** 305–310.

CHEN, C.-H., K. K. OISHI, B. KLOECKENER-GRUISSEN and M. FREELING, 1987 Organ-specific expression of maize *Adh1* is altered after a *Mu* transposon insertion. Genetics **116:** 469–477.

CRESSE, A. C., S. H. HULBERT, W. E. BROWN, J. R. LUCUS and J. L. BENNETZEN, 1994 *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. Genetics **140:** 315–324.

DAVIS, L. G., M. D. DIBNER and J. F. BATTEY, 1986 *Basic Methods in Molecular Biology*. Elsevier Scientific Publishers, New York.

DEVEREUX, J., P. HAEBERLIE and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DOGBO, O., A. LAFERRIERE, A. D'HARLINGUE and B. CAMARA, 1988 Carotenoid biosynthesis: isolation and characterization of a bifunctional enzyme catalyzing the synthesis of phytoene. Proc. Natl. Acad. Sci. USA **85:** 7054–7058.

FROHMAN, M. A., M. K. DUSH and G. R. MARTIN, 1988 Rapid production of full-length cDNA from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

GIULIANO, G., G. E. BARTLEY and P. A. SCOLNIK, 1993 Regulation of carotenoid biosynthesis during tomato development. Plant Cell **5:** 379–387.

GOODWIN, T. W., 1971 Biosynthesis, pp. 577–636 in *Carotenoids*, edited by O. ISLER. Birkhauser Verlag, Basel.

HELENTJARIS, T., D. L. WEBER and S. WRIGHT, 1988 Identification of the genomic location of duplicated nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics **118:** 353–363.

JIN, Y.-K., and J. L. BENNETZEN, 1994 Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the *Bs1* retroelement of maize. Plant Cell **6:** 1177–1186.

KARVOUNI, Z., I. JOHN, J. E. TAYLOR, C. F. WATSON, A. J. TURNER *et al.*, 1995 Isolation and characterization of a melon cDNA clone encoding phytoene synthase. Plant Mol. Biol. **27:** 1153–1162.

LOH, E. Y., J. F. ELLIOTT, S. CWIRLA, L. L. LANIER and M. M. DAVIS, 1989 Polymerase chain reaction with single-sided specificity: Analysis of T cell receptor δ chain. Science **243:** 217–220.

LUEHRSEN, K. R., and V. WALBOT, 1990 Insertion of *Mu1* elements in the first intron of the *Adh1-S* gene of maize results in novel RNA processing events. Plant Cell **2:** 1225–1238.

MANGELSDORF, P. C., and G. S. FRAPS, 1931 A direct quantitative relationship between vitamin A in corn and the number of genes for yellow pigmentation. Science **73:** 241–242.

MASON, P. J., and J. G. WILLIAMS, 1987 Hybridization in the analysis of recombinant DNA, pp. 113–137 in *Nucleic Acid Hybridization A Practical Approach*, edited by B. D. HAMES and S. J. HIGGINS. IRL Press, Washington, DC.

MCCARTY, D. R., 1986 A simple method for extraction of RNA from maize tissue. Maize Gen. Coop. Newslett. **60:** 61.

MISAWA, N., M. NAKAGAWA, K. KOBAYASHI, S. YAMANO, Y. IZAWA *et al.*, 1990 Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J. Bacteriol. **172:** 6704–6712.

MORGANTE, M., and A. M. OLIVIERI, 1993 PCR-amplifed microsatellites as markers in plant genetics. Plant J. **3:** 175–182.

PARRY, A. D., and R. HORGAN, 1991 Carotenoids and abscisic acid (ABA) biosynthesis in higher plants. Physiol. Plant. **82:** 320–326.

PAZ-ARES, J., D. GHOSAL, U. WIENAND, P. A. PETERSON and H. SAEDLER, 1987 The regulatory *c1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and structural similarities to transcriptional activators. EMBO J. **6:** 3553–3558.

PHELPS, T. L., and B. BUCKNER, 1995 Allele-specific length polymorphisms associated with microsatellite repeat variation within the *y1* gene. Maize Gen. Coop. Newslett. **69:** 84–85.

RALSTON, E. J., J. J. ENGLISH and H. K. DOONER, 1988 Sequence of three *bronze* alleles of maize and correlation with the genetic fine structure. Genetics **119:** 185–197.

RANDALL, S. K., M. S. MARSHALL and D. N. CROWELL, 1993 Protein isoprenylation in suspension-cultured tobacco cells. Plant Cell **5:** 433–442.

RAY, J., C. BIRD, M. MAUNDERS, D. GRIERSON and W. SCHUCH, 1987 Sequence of pTOM5, a ripening related cDNA from tomato. Nucleic Acids Res. **15:** 10587.

RAY, J., P. MOUREAU, C. BIRD, A. BIRD, D. GRIERSON *et al.*, 1992 Cloning and characterization of a gene involved in phytoene synthesis from tomato. Plant Mol. Biol. **19:** 401–404.

ROBERTSON, D. S., 1975 Survey of the albino and white-endosperm mutants of maize. Their phenotype and gene symbols. J. Hered. **66:** 67–74.

ROBERTSON, D. S., 1985 Differential activity of the maize mutator *Mu* at different loci and in different cell lineages. Mol. Gen. Genet. **200:** 9–13.

ROBERTSON, D. S., and I. C. ANDERSON, 1961 Temperature sensitive alleles of the *y1* locus in maize. J. Hered. **52:** 53–60.

ROMER, S., P. HOGUENEY, F. BOUVIER, B. CAMARA and M. KUNTZ, 1993 Expression of the genes encoding the early carotenoid biosynthetic enzymes in *Capsicum annum*. Biochem. Biophys. Res. Commun. **196:** 1414–1421.

SANGER, F., S. NICKLESON and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHMIDHAUSER, T. J., F.-R. LAUTER, M. SCHUMACHER, W. ZHOU, V. E. A. RUSSO *et al.*, 1994 Characterization of *al-2*, the phytoene synthase gene of *Neurospora crassa*. J. Biol. Chem. **269:** 12060–12066.

SCOLNIK, P. A., and G. E. BARTLEY, 1994 Nucleotide sequence of an *Arabidopsis* cDNA for phytoene synthase. Plant Physiol. **104:** 1471–1472.

SOLOMON, L. R., and P. A. RUBENSTEIN, 1987 Studies on the role of actin N-methylhistidine using oligonucleotide-directed site-specific mutagenesis. J. Biol. Chem. **262:** 11382–11388.

STROMMER, J. N., and D. F. ORTIZ, 1989 *Mu1*-induced mutant alleles of maize exhibit background-dependent changes in expression and RNA processing. Dev. Genet. **10:** 452–459.

THOMAS, P. S., 1980 Hybridization of denatured RNA and small DNA fragments to nitrocellulose. Proc. Natl. Acad. Sci. USA **77:** 5201–5205.

Communicating editor: W. F. SHERIDAN